

Mapping Quantitative Trait Loci in Experimental Populations

by

Gary A. Churchill

Department of Plant Breeding and Biometry
Cornell University,
Ithaca, NY 14853

Rebecca W. Doerge

Department of Plant Breeding and Biometry
Cornell University,
Ithaca, NY 14853

BU-1238-M

May 1994

Mapping Quantitative Trait Loci in Experimental Populations

Gary A. Churchill and Rebecca W. Doerge

Department of Plant Breeding and Biometry
Cornell University, Ithaca, NY 14853, U.S.A.

SUMMARY

A general framework for mapping quantitative trait loci (QTL) in experimental populations is presented. The approach is based on an augmented data formulation of the likelihood. The observable data consist of trait values and marker phenotypes on each of n plants. In the augmented data formulation, the QTL and marker genotypes for each individual are also given. Inferential problems to be addressed are the detection of genetic factors, QTL, affecting the distribution of the trait, the location of QTL relative to DNA markers and the estimation of QTL effects. An EM algorithm for computing observed data maximum likelihood estimates of the model parameters is described. This approach is applied to a backcross population with a single segregating QTL and to an intercross population with two segregating QTL. In the second example, dominance and epistatic interactions among the QTL alleles are introduced.

1. Introduction

Mapping genes that control quantitative traits is an important problem in plant genetics. In this paper, we examine the problem of model based inferences of QTL effects in experimental populations. The statistical problem of QTL mapping can be viewed as having three components. First is the *detection* of genetic factors, QTL, segregating in the population that affect the distribution of the trait of interest. Second is the *location* of QTL relative to a set of DNA markers. Third is the *estimation* of the QTL effects and their interactions. These problems are not entirely separate, but the distinction will help to clarify the inferential procedures used in QTL mapping.

The QTL mapping problem has received a significant amount of attention over the past 10 years. We list a few key references but will not attempt to make a complete survey of the literature. Some of the earliest work on likelihood methods for the mapping and analysis of quantitative traits is due to Weller (1986, 1987). The method of interval mapping, as described by Lander and Botstein (1989) has been the most widely used approach to quantitative trait mapping. A general formulation of the QTL inference problems closely related to the present work is described by Jansen (1993a). New ideas and

approaches continue to be developed, for example, see Jansen (1993b), Jansen and Stam (1994) and Zeng (1993, 1994).

The remainder of this paper is organized as follows. In section 2, we examine a general model of QTL effects and identify components of an augmented data likelihood that must be specified in any particular situation. The problems of detection and location of QTL effects are briefly discussed. An EM algorithm approach to the estimation and location of QTL effects is described. In section 3, the likelihood and EM algorithm are developed for two examples. We first consider a backcross population with a single segregating QTL and discuss inference for a single marker, then for a pair of markers spanning the QTL and finally extend this result to a full map analysis. Next we consider an F2 population segregating for two unlinked QTL. A particular form of genetic interaction is described. In the final section, we discuss some problem areas where further statistical work is needed on the QTL mapping problem.

2. Modeling QTL

2.1 QTL and Mixtures

Consider a population of plants indexed by $i = 1, \dots, n$. Let Y_i be the quantitative trait value of the i th plant, *e.g.* its height. In the simplest setting, we observe a single DNA marker with two alleles,

$$M_i = \begin{cases} 0 & \text{absent} \\ 1 & \text{present.} \end{cases}$$

We assume that there is a single QTL, also with two alleles,

$$Q_i = \begin{cases} 0 & \text{low} \\ 1 & \text{high} \end{cases}$$

segregating in the population.

The effect of the QTL on the trait is typically modeled as a location shift although more general models can be considered. Within the class of individuals defined by $Q_i = 0$, the trait value is a random variable with density $p_{Y|Q}(y, 0) = f(y)$ and within the class of individuals defined by $Q_i = 1$, the trait values are distributed with density $p_{Y|Q}(y, 1) = f(y - \Delta)$. Thus the marginal distribution of the trait values will be a mixture with the mixing proportions determined by the frequency of the QTL alleles in the population.

The allelic state of the QTL cannot be directly observed. However, we can observe the marker class M_i of each plant. If the QTL and the marker are linked and we let r denote the recombination fraction, *i.e.* $r = \Pr(Q_i \neq M_i)$,

the conditional densities of the trait value within marker classes are also mixtures

$$p_{Y|M}(y, m) = r^m(1 - r)^{1-m}f(y) + r^{1-m}(1 - r)^mf(y - \Delta). \quad (1)$$

Note that the means of the conditional densities $p_{Y|M}(y, 0)$ and $p_{Y|M}(y, 1)$ will differ by $(1 - 2r)\Delta$. This location change is the key to QTL detection.

2.2 Testing for QTL Effects

There are three hypotheses relevant to the QTL detection problem (Knott and Haley 1992) these being

$H_0^1 : \Delta = 0$; no QTL is present,

$H_0^2 : r = 1/2, \Delta > 0$; a QTL is present but is not linked to the marker,

$H_A : r < 1/2, \Delta > 0$; a QTL is present and is linked to the marker.

For a given density $f()$, likelihood ratio tests can be constructed to compare H_A to either of the two null hypotheses. However, the asymptotic properties of these tests present some difficulties (*e.g.* Hartigan 1985; Gosh and Sen 1985). In practice, a t -test is commonly used to detect the location shift and although the assumptions required for the t -test are not satisfied here, it behaves quite well in simulations (Doerge, 1993). Churchill and Doerge (1994) have recommended the use of a permutation test that is valid for any continuous density $f()$ and provides an unbiased test under either of the null hypotheses.

The detection and location problems are closely connected. In a typical mapping experiment, hundreds of markers may be available and tests will be carried out at each marker. If the markers are organized into a map, tests may also be carried out by assuming the QTL is present in intervals between markers. A QTL effect is detected if association with any of the markers is discovered. The location of the QTL will be inferred by identifying those markers that are most strongly associated with the trait. There is a multiple testing problem here. An asymptotic argument for determining valid experimentwise critical values has been presented by Lander and Botstein (1989, 1994). However, the permutation test of Churchill and Doerge (1994) can also be used to obtain experimentwise critical values without appeal to asymptotic arguments.

2.3 Augmented Data Likelihood

Again consider n plants with observed trait values Y_i and marker phenotypes M_i . The “missing” data for the QTL mapping problem are Q_i , the QTL

genotype and G_i the marker genotype. In this section, both the QTL and marker genotypes may consist of one or more loci. The distinction between marker phenotypes and marker genotypes is necessary in cases where there are dominant markers or in experimental crosses, such as an intercross, where the relative phases between loci may be unknown.

With this notation, we can express the likelihood as

$$\Pr(\mathbf{Y}, \mathbf{M}) = \prod_{i=1}^n \Pr(Y_i, M_i) \quad (2)$$

$$= \prod_{i=1}^n \sum_{G_i} \sum_{Q_i} \Pr(Y_i, M_i, G_i, Q_i) \quad (3)$$

$$= \prod_{i=1}^n \sum_{G_i} \sum_{Q_i} \Pr(Y_i, M_i | G_i, Q_i) \Pr(G_i, Q_i) \quad (4)$$

$$= \prod_{i=1}^n \sum_{G_i} \sum_{Q_i} \Pr(Y_i | G_i, Q_i) \Pr(M_i | G_i, Q_i) \Pr(G_i, Q_i) \quad (5)$$

$$= \prod_{i=1}^n \sum_{G_i} \sum_{Q_i} \Pr(Y_i | Q_i) \Pr(M_i | G_i) \Pr(G_i, Q_i). \quad (6)$$

Dependence on a vector θ of model parameters is implicit throughout. The product in (2) follows from the assumption that plants are independent. The likelihood is expressed as a mixture in (3) and factored using the definition of conditional probability in (4). Conditional independence of the trait value and the marker phenotype is assumed in (5) which seems reasonable in most cases. However, the conditional independence (of Y_i and G_i given Q_i) assumption used to derive (6) should be examined carefully. If there are additional loci in the genome that affect the trait value distribution and if they are linked to the marker(s), this conditional independence will not hold. We will proceed by making this assumption but note that it may be worthwhile to pursue models which do not.

Using this factorization of the likelihood, there are 3 components of the model that must be specified.

1. The conditional distribution of the trait value given the QTL genotype $\Pr(Y_i | Q_i)$ may be taken to be any exponential family distribution. At present, most QTL analyses are carried out by assuming normal distributions. In many cases, other distributions (*e.g.* exponential distributions for lifetimes or Poisson distributions for counts) may be more appropriate.

2. The conditional distribution of marker phenotypes given marker genotypes $\Pr(M_i | G_i)$ will be multinomial on the marker phenotype classes but will typically be degenerate, *i.e.* all class probabilities equal to 0 except one equal to 1. An interesting case where $\Pr(M_i | G_i)$ will not be degenerate is if marker phenotyping errors are introduced into the model (Lincoln and Lander, 1992).
3. Segregation and linkage between the QTL and the marker(s) $\Pr(G_i, Q_i)$ will again be multinomial. The number of categories and their probabilities will be determined by the number of markers and the experimental design. Segregation distortion or crossover interference can be introduced into the model by modifying this piece of the likelihood.

The augmented data likelihood

$$\prod_{i=1}^n \Pr(Y_i, M_i, G_i, Q_i) = \prod_{i=1}^n \Pr(Y_i | Q_i) \Pr(M_i | G_i) \Pr(G_i, Q_i) \quad (7)$$

will be an exponential family distribution.

2.4 An Estimation Algorithm

Maximum likelihood parameter estimates can be obtained by the following algorithm, a special case of the EM algorithm (Dempster *et al.*, 1977). Starting with an initial estimate of the parameter $\theta^{(0)}$, iterate the following two steps.

E-step: Compute $E(Q_i, G_i | Y_i, M_i)$ using the current estimate $\theta^{(p)}$ after the p th iteration. The genotypes G_i and Q_i can be represented as indicator vectors, thus the desired expectations follow directly from the conditional probability density

$$\begin{aligned} \Pr(Q_i, G_i | Y_i, M_i) &\propto \Pr(Y_i, M_i | Q_i, G_i) \Pr(Q_i, G_i) \\ &= \Pr(Y_i | Q_i) \Pr(M_i | G_i) \Pr(Q_i, G_i) \end{aligned} \quad (8)$$

The constant of proportionality $\sum_{G_i} \sum_{Q_i} \Pr(Y_i, M_i | Q_i, G_i) \Pr(Q_i, G_i)$, will be tractable for single markers or small sets of marker. For large sets of markers more elaborate algorithms may be required (Lander and Green, 1987).

M-step: Obtain new parameter estimates $\theta^{(p+1)}$ replacing Q_i and G_i by their conditional expectations in the augmented data likelihood. For exponential family distributions, the estimation becomes a standard problem in generalized linear models (McCullagh and Nelder, 1989).

The E-step and the M-step are iterated until convergence is obtained in the parameter estimates. A number of well placed starting values should be tested to ensure that convergence to a global maximum has been obtained.

3. Examples

3.1 Single QTL in a Backcross Population

We first consider the problem of estimating of the recombination fraction between a single marker locus A and a quantitative trait locus Q in a backcross population. The complete genotypic state of a backcross individual i is specified by two indicator functions for the presence/absence of the non-recurrent parental allele,

$$Q_i = \begin{cases} 0 & \text{absent} \\ 1 & \text{present} \end{cases} \quad \text{and} \quad A_i = \begin{cases} 0 & \text{absent} \\ 1 & \text{present} \end{cases}.$$

The marker phenotype and marker genotype are identical in this design so the component $\Pr(M_i | G_i)$ can be dropped from the model. Let r denote the probability of a recombination between Q and A per chromosome per generation and assume regular Mendelian segregation. The linkage and segregation component of the model is specified by enumerating the four possible genotype configurations and counting recombination events. Thus

$$\begin{aligned} \Pr(A = 1, Q = 1) &= \Pr(A = 0, Q = 0) = (1 - r)/2 \\ \Pr(A = 1, Q = 0) &= \Pr(A = 0, Q = 1) = r/2. \end{aligned} \quad (9)$$

We will assume that the trait distributions $\Pr(Y_i | Q_i)$ are normal within each QTL genotype class and that the classes have a common variance. Thus

$$Y_i \sim N(\mu_i, \sigma^2) \quad (10)$$

where

$$\mu_i = \begin{cases} \nu_0 & \text{if } Q_i = 0 \\ \nu_1 & \text{if } Q_i = 1. \end{cases} \quad (11)$$

For identifiability, we assume $\nu_0 \neq \nu_1$.

This is the “standard” QTL model for a backcross population. Some generalizations are immediately available to us in the present framework. First, the assumption of common variance σ^2 can be relaxed with only minor changes to the analysis below. This is important as in practice both the mean and the variance of a trait may be affected by the QTL. Second, the assumption of a normal distribution within genotype classes can be replaced with any distribution. Modifications to the analysis below will be relatively minor provided we stay within the class of exponential family distributions. A number of other generalizations are possible. For example non-Mendelian segregation could be introduced as an additional parameter in the genotype class distribution, replacing the factor $1/2$ in (9).

M-step: If the QTL states $Q_i = q_i$ were known for each plant, we could obtain simple direct estimates of all model parameters by maximizing the augmented data likelihood,

$$\Pr(\mathbf{Y}, \mathbf{M}, \mathbf{Q}) = \theta^x (1 - \theta)^{n-x} \prod_{i=1}^n \left[\frac{q_i}{\sigma} \phi\left(\frac{y_i - \nu_1}{\sigma}\right) + \frac{1 - q_i}{\sigma} \phi\left(\frac{y_i - \nu_0}{\sigma}\right) \right], \quad (12)$$

where

$$x = \sum_{i=1}^n q_i(1 - a_i) + (1 - q_i)a_i, \quad (13)$$

a_i is the observed marker state, y_i is the observed trait value and $\phi()$ is the standard normal density function.

The augmented data maximum likelihood estimators are

$$\begin{aligned} \hat{r} &= x/n \\ \hat{\nu}_0 &= \sum_{i=1}^n (1 - q_i)y_i / \sum_{i=1}^n (1 - q_i) \\ \hat{\nu}_1 &= \sum_{i=1}^n q_i y_i / \sum_{i=1}^n q_i \\ \hat{\sigma}^2 &= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / n \end{aligned} \quad (14)$$

where $\hat{\mu}_i = (1 - q_i)\hat{\nu}_0 + q_i\hat{\nu}_1$.

E-step: In this problem, because the marker phenotype and genotype are identical, we compute the conditional expectation of the QTL genotype state given the observed phenotype, marker genotype and the current estimate of the model parameters,

$$E(Q_i | Y_i = y, A_i = a) = \frac{r^{1-a}(1-r)^a \phi\left(\frac{y-\nu_1}{\sigma}\right)}{r^{1-a}(1-r)^a \phi\left(\frac{y-\nu_1}{\sigma}\right) + r^a(1-r)^{1-a} \phi\left(\frac{y-\nu_0}{\sigma}\right)}. \quad (15)$$

Two Markers: We can extend this model to the case of two markers

$$A_i = \begin{cases} 0 & \text{absent} \\ 1 & \text{present} \end{cases} \quad \text{and} \quad B_i = \begin{cases} 0 & \text{absent} \\ 1 & \text{present} \end{cases}.$$

There are three possible arrangements of two markers and one QTL when all three are linked. However it is only necessary to consider the case A-Q-B where the QTL is located in the interval between the two markers. This is because, with the assumption of independence between recombination events

in different intervals, the cases Q–A–B and A–B–Q reduce to the single marker problems Q–A and B–Q respectively. Let r_A be the recombination fraction between the QTL and marker A and let r_B be the recombination fraction between the QTL and marker B . The joint distribution of genotypes at A , Q and B is

$$\begin{aligned}\Pr(Q_i = 0, A_i = 0, B_i = 0) &= \Pr(Q_i = 1, A_i = 1, B_i = 1) = \frac{1}{2}(1 - r_A)(1 - r_B) \\ \Pr(Q_i = 0, A_i = 0, B_i = 1) &= \Pr(Q_i = 1, A_i = 1, B_i = 0) = \frac{1}{2}(1 - r_A)r_B \\ \Pr(Q_i = 0, A_i = 1, B_i = 0) &= \Pr(Q_i = 1, A_i = 0, B_i = 1) = \frac{1}{2}r_A(1 - r_B) \\ \Pr(Q_i = 0, A_i = 1, B_i = 1) &= \Pr(Q_i = 1, A_i = 0, B_i = 0) = \frac{1}{2}r_Ar_B.\end{aligned}$$

M-step: If the QTL genotypes are known we can write the augmented data likelihood in a form similar to (12) and obtain maximum likelihood estimates from the augmented data:

$$\hat{r}_A = \frac{x_A + x_Q}{n} \quad \text{and} \quad \hat{r}_B = \frac{x_B + x_Q}{n} \quad (16)$$

where

$$\begin{aligned}x_A &= \sum_{i=1}^n a_i(1 - b_i)(1 - q_i) + (1 - a_i)b_iq_i \\ x_B &= \sum_{i=1}^n (1 - a_i)b_i(1 - q_i) + a_i(1 - b_i)q_i \\ x_Q &= \sum_{i=1}^n (1 - a_i)(1 - b_i)q_i + a_ib_i(1 - q_i).\end{aligned} \quad (17)$$

maximum likelihood estimates of ν_0 , ν_1 and σ^2 are obtained as above (14).

E-step: The conditional expectations follow from

$$\begin{aligned}\Pr(Q_i | Y_i, A_i, B_i) &\propto \Pr(Q_i)\Pr(Y_i, A_i, B_i | Q_i) \\ &= \Pr(Q_i)\Pr(Y_i | Q_i)\Pr(A_i | Q_i)\Pr(B_i | Q_i)\end{aligned} \quad (18)$$

where we have assumed conditional independence of A , B and Y given Q . These are

$$\begin{aligned}E(Q_i | Y_i = y, A_i = a, B_i = b) &= \\ &= \frac{r_A^{1-a}(1 - r_A)^a r_B^{1-b}(1 - r_B)^b \phi\left(\frac{y - \nu_1}{\sigma}\right)}{r_A^{1-a}(1 - r_A)^a r_B^{1-b}(1 - r_B)^b \phi\left(\frac{y - \nu_1}{\sigma}\right) + r_A^a(1 - r_A)^{1-a} r_B^b(1 - r_B)^{1-b} \phi\left(\frac{y - \nu_0}{\sigma}\right)}\end{aligned} \quad (19)$$

Multiple Markers: Consider a map with m markers in a known order. For each interval in the map we can compute the maximized log-likelihood assuming a QTL is located in that interval using the EM algorithm above. This approach is valid if we assume one QTL and independence of recombination events within each plant. Thus, for the k th interval we obtain a maximized likelihood $L_k(\hat{\theta})$. Likelihoods for distinct intervals can be compared and the interval with the highest likelihood is a maximum likelihood estimate of the QTL location.

3.2 Two QTL in an Intercross Population

We consider a trait Y with distribution determined by two QTL. In the augmented data setting where the QTL genotype is known and the conditional trait distributions are normal with common variance, the estimation problem is equivalent to the standard two-way analysis of variance. In an intercross population there are three possible genotypes at each locus. For linked loci, we must also consider the relative phases of these loci. For unlinked QTL, the possible genotypes can be represented as a pair of indicator vectors $Q_i = (Q_{i1}, Q_{i2})$,

$$Q_{ij} = \begin{cases} (1, 0, 0)^T & \text{homozygous 11} \\ (0, 1, 0)^T & \text{heterozygous 12} \\ (0, 0, 1)^T & \text{homozygous 22} \end{cases}$$

for $i = 1, \dots, n$, $j = 1, 2$. The two QTL in this system are assumed to be unlinked, thus there are nine possible values for Q_i .

The model presented here was motivated by work on the expression of acylsugars in tomatoes derived from an intercross between a wild species and a cultivar (Mutschler and Shapiro, 1994). The observed phenotype for each plant consists of a bivariate observation $Y_i = (Y_{i1}, Y_{i2})$, where Y_{i1} is the total acylsugar detected in a standard assay and Y_{i2} is the proportion of glucose acylsugar among the total acylsugar. Data from a population of 196 plants are shown in figure 1.

The following genetic model is proposed as a working hypothesis. There are two major QTL in this system plus other modifiers that may be genetic or environmental. We assume that the two QTL are unlinked and that any additional genetic modifiers are unlinked to the two QTL. The first QTL affects the level of acylsugar production. The high production allele is dominant to low production allele. The second QTL affects the proportion of glucose among the high level producers. It has no effect on the low level producers. The low glucose allele is dominant to the high glucose allele. The genetic model is summarized in the following table.

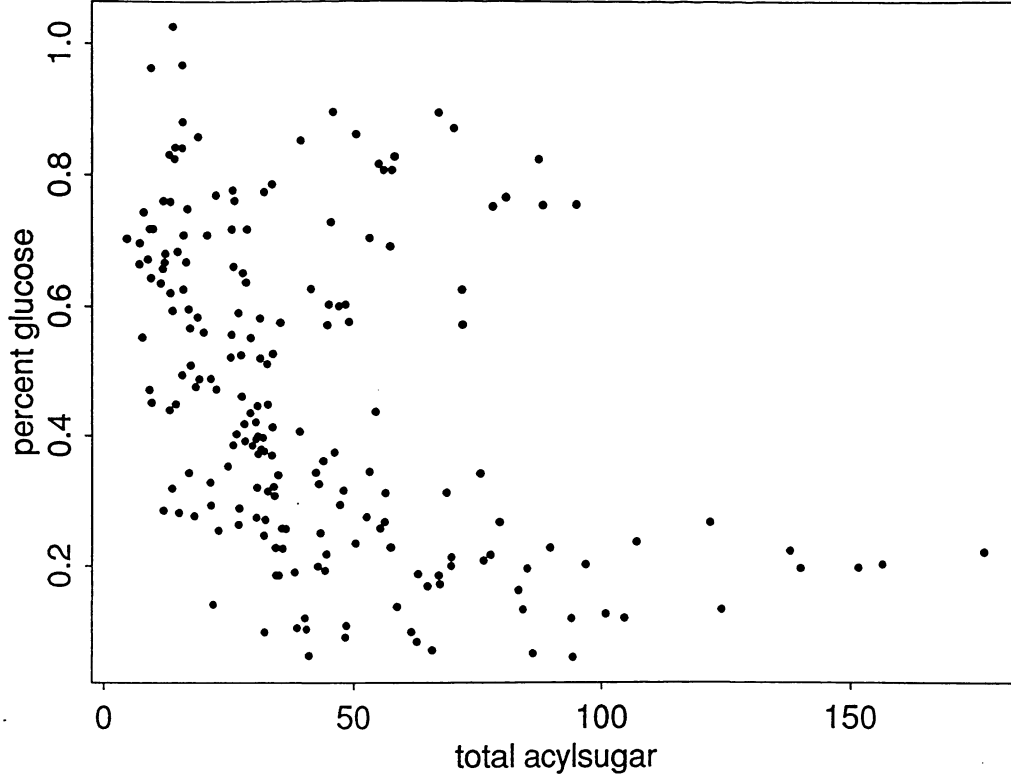


Figure 1: Acylsugar Trait Data

		Q_1		
		11	12	22
Q_2	11	I	II	II
	12	I	II	II
	22	I	III	III

The conditional trait distributions $\Pr(Y_i | Q_i)$ are

$$\begin{aligned}
 Y|\text{group I} &\sim N_2(\mu_1, \Sigma_1) \\
 Y|\text{group II} &\sim N_2(\mu_2, \Sigma_2) \\
 Y|\text{group III} &\sim N_2(\mu_3, \Sigma_3)
 \end{aligned} \tag{20}$$

where

$$\Sigma_i = \begin{bmatrix} \sigma_{i1} & 0 \\ 0 & \sigma_{i2} \end{bmatrix}.$$

We assume that we have located two markers, $M_i = (M_{1i}, M_{2i})$, such that the first is linked to Q_{i1} and the second is linked to Q_{i2} . The distributions $\Pr(M_i | G_i)$ and $\Pr(G_i, Q_i)$ follow from standard intercross genetics. There

are 100 (phase-known) genotypes to enumerate, ten at each marker-QTL pair in all combinations. The enumeration is straightforward, and is not shown here.

Using the augmented data formulation presented here, we have developed an EM algorithm to estimate the model parameters. We are in the process of examining all pairs of markers scored on this population to locate the best markers for estimation of the QTL effects.

4. Discussion

We have described a general approach to modeling the effects of QTL and developed inference procedures based on this model. Model based methods are always subject to the criticism that the models are not correct. We acknowledge that some of the assumptions required here oversimplify the reality of quantitative genetics. However, with this formulation it is clear where the various assumptions enter into the analysis and which components of the model should be modified. The modeling approach requires the researcher to specify the number of QTL involved in a system and the nature of any interactions between multiple QTL. We do not consider this to be a disadvantage. The model, and hence the inference procedures, are specifically directed to the problem at hand. It is possible with this approach to compare various alternative models to test hypotheses about the genetic system and also to make checks on the adequacy of the model.

The QTL mapping problem presents new and interesting statistical challenges, many of which remain unsolved. Some open problems of practical importance include estimation of the number of (major) QTL in a genetic system, development of efficient algorithms for localizing multiple QTL and modeling genetic interactions. We hope that this introduction will attract more statisticians to work on these problems.

ACKNOWLEDGEMENTS

The authors are grateful to Martha Mutschler for providing the data used in the second example. This work was supported in part by grants from the U.S. Department of Agriculture and the U.S. Department of Energy.

REFERENCES

- Churchill, G.A. and Doerge, R.W. (1994). Empirical thresholds values for quantitative trait mapping. Technical Report, Biometrics Unit, Cornell University.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–22.
- Doerge, R.W. (1993). *Statistical Methods for Locating Quantitative Trait Loci with Molecular Markers*. Ph.D. Dissertation, Department of Statistics, North Carolina State University.
- Ghosh, J.K. and Sen, P.K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol.II, 789–806.
- Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal distributions. *Proceedings of the Berkeley Conference*, Vol.II, 807–810.
- Jansen, R.C. (1993a). A general mixture model for mapping quantitative trait loci by using molecular markers. *Theoretical and Applied Genetics* **85**, 252–260.
- Jansen, R.C. (1993b). Interval Mapping of multiple quantitative trait loci. *Theoretical and Applied Genetics* **79**, 583–592.
- Jansen, R.C. and Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.
- Knott, S.A. and Haley, C.S. (1992). Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. *Genetical Research* **60**, 139–151.
- Lander, E.S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.
- Lander, E.S. and Botstein, D. (1994). Corrigendum. *Genetics* **36**, 705.

- Lander, E. and Green, P. (1987). Construction of multilocus genetic maps in humans. *Proceedings of the National Academy of Sciences, USA* **84**, 2363–2367.
- Lincoln, S. and Lander, E. (1992). Systematic detection of errors in genetic linkage data. *Genomics* **14**, 604–610.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, second ed. London: Chapman and Hall.
- Mutschler, M.A., and Shapiro, Y. (1994). *Biochemical Systematics and Ecology*, in press.
- Weller, J.I. (1986). Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* **42**, 627–640.
- Weller, J.I. (1987). Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using appropriate maximum likelihood methods. *Heredity* **59**, 413–421.
- Zeng, Z-B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences, USA* **90**, 10972–10976.
- Zeng, Z-B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.